

## **Design optimization for clinical trials in early stage manifest Huntington's disease**

**Chris Frost MA, DipStat<sup>a</sup>, Amy Mulick MSc<sup>a</sup>, Rachael I. Scahill PhD<sup>b</sup>, Gail Owen PhD<sup>b</sup>, Elizabeth Aylward PhD<sup>c</sup>, Blair R. Leavitt MDCM<sup>d</sup>, Alexandra Durr MD<sup>e</sup>, Raymund A. C. Roos MD, PhD<sup>f</sup>, Beth Borowsky PhD<sup>g</sup>, Julie C. Stout PhD<sup>h</sup>, Ralf Reilmann MD<sup>i</sup>, Douglas R. Langbehn MD, PhD<sup>j</sup>, Sarah J. Tabrizi MD, PhD<sup>b</sup>, Cristina Sampaio MD, PhD<sup>k</sup> and the TRACK-HD Investigators**

- <sup>a</sup> Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK
- <sup>b</sup> UCL Institute of Neurology, University College London, Queen Square, London, WC1N 3BG, UK
- <sup>c</sup> Center for Integrative Brain Research, Seattle Children's Research Institute Seattle, WA 98195-7115
- <sup>d</sup> Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada
- <sup>e</sup> ICM - Institut du Cerveau et de la Moelle Epinière, INSERM U1127, CNRS UMR7225, Sorbonne Universités – UPMC Université Paris VI UMR\_S1127 and APHP, Genetic department, Pitié-Salpêtrière University Hospital, Paris, France
- <sup>f</sup> Department of Neurology, Leiden University Medical Centre, 2300RC Leiden, The Netherlands
- <sup>g</sup> CHDI Management Inc/CHDI Foundation, Princeton, USA; and Clinical Development, Neurodegenerative Diseases, Teva Pharmaceuticals, Malvern, USA
- <sup>h</sup> School of Psychological Sciences, Monash University, Melbourne, Victoria, Australia
- <sup>i</sup> George Huntington Institute, Technology Park Muenster, Johann-Krane-Weg 27, 48149 Muenster, Institute for Clinical Radiology, University of Muenster, Albert Schweitzer Campus 1 – Building A1, 48149 Muenster, Germany and Department of Neurodegenerative Diseases and Hertie-Institute for Clinical Brain Research, University of Tuebingen, Hoppe-Seyler Str. 3, 72076 Tuebingen, Germany
- <sup>j</sup> Department of Psychiatry, University of Iowa, Iowa City, Iowa 52242, USA
- <sup>k</sup> CHDI Management Inc/CHDI Foundation, Princeton, USA

Design optimization for clinical trials in early stage manifest Huntington's Disease

Corresponding author: Chris Frost, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK. +44 (0) 207 927 2242. [chris.frost@lshtm.ac.uk](mailto:chris.frost@lshtm.ac.uk)

Word count: 3694

**Abstract:**

Objectives: To inform the design of randomized clinical trials in early stage manifest Huntington's Disease through analysis of longitudinal data from TRACK-Huntington's Disease (TRACK-HD), a multi-centre observational study.

Methods: We compute sample sizes required for trials with candidate clinical, functional and imaging outcomes, whose aims are to reduce rates of change. The calculations use a two-stage approach: first using linear mixed models to estimate mean rates of change and components of variability from TRACK-HD data, and second using these to predict sample sizes for a range of trial designs.

Results: For each outcome the primary drivers of required sample size were the anticipated treatment effect and the duration of treatment. Extending durations from one to two years yielded large sample size reductions. Including interim visits, and incorporating stratified randomization on predictors of outcome together with covariate adjustment, gave more modest, but non-trivial, benefits. Caudate atrophy expressed as a percentage of its baseline, gave smallest sample sizes.

Discussion: Here we consider potential required sample sizes for clinical trials estimated from naturalistic observation of longitudinal change. Choice amongst outcome measures for a trial must additionally consider their relevance to patients, and the expected effect of the treatment under study. For all outcomes considered our results provide compelling arguments for two-year trials, and we also demonstrate benefits of incorporating stratified randomization coupled with covariate adjustment, particularly for trials with caudate atrophy as the primary outcome. The benefits of enrichment are more debatable, with statistical benefits offset by potential recruitment difficulties and reduced generalizability.

**Background:**

Other than interventions targeting chorea, clinical trials testing therapeutics in Huntington's disease (HD) have not provided evidence of efficacy<sup>1</sup>. There are several potential reasons for these failures, including a real absence of effect, and poor choice of design and outcomes. New classes of drugs such as huntingtin (HTT) lowering therapies are now entering clinical trials<sup>2</sup>. Thus, there is need for assessment of trial designs to ensure that future trials are optimally designed.

Choosing the most appropriate trial design and outcome measure involves consideration of many factors, including relevance of outcomes to patients, and treatment target. Statistical efficiency, measured by the sample size necessary to detect clinically meaningful treatment effects is also key, and is our focus here.

Currently we lack data from HD clinical trials illustrating how potential outcomes behave in response to an effective intervention. In such situations trial design may be informed by other data, such as that from longitudinal cohort studies, with decisions informed by modelling and simulation<sup>3, 4</sup>. Here we exploit the TRACK-HD longitudinal study<sup>5-8</sup>. We consider interventions aiming to alter rates of disease progression. Accordingly, outcome measures are subject-specific rates of change, rather than absolute levels, of structure and function. We model the longitudinal TRACK-HD data using linear mixed models<sup>9</sup> that explicitly estimate between- and within-subject components of variance, thereby allowing prediction of statistical efficiency for different trial designs. Specifically, we consider: a) choice of outcome; b) duration of follow-up and assessment periodicity; c) stratified randomization using baseline measures potentially predictive of outcome, coupled with statistical adjustment for these measures; (d) enrichment using such baseline measures; e) plausible dropout patterns.

## **Methods:**

In TRACK-HD<sup>5-8</sup> HD patients with early stage disease, pre-manifest gene carriers and controls were assessed at baseline, 12, 24, and 36 months at sites in London, Paris, Leiden and Vancouver. Here we only consider the early-stage HD patients and controls. Using the baseline Unified Huntington's Disease Rating Scale (UHDRS) Total Functional Capacity (UHDRS-TFC) score<sup>10</sup>, early-HD patients were designated either stage 1 (UHDRS-TFC 11-13) or stage 2 (UHDRS-TFC 7-10).

We selected six potential outcomes based on their high frequency of use in HD clinical research<sup>11</sup>: the UHDRS-TFC, the UHDRS Total Motor Score (UHDRS-TMS)<sup>12</sup>, the Symbol Digit Modalities Test (SDMT) score, the speeded tapping inter-onset interval standard deviation (IOI-SD) of the quantitative motor assessments for the non-dominant hand<sup>13</sup>, whole brain atrophy and caudate atrophy. Changes in whole brain and caudate volumes were measured using the boundary shift integral (BSI) methodology<sup>14</sup> and the resultant rates of change numerically expressed in three ways: i) ml per year, which represents raw volume loss<sup>15</sup> (mL), ii) annualized percentage of the volume of the baseline region of interest (%base)<sup>6</sup> and iii) annualized percentage of total intracranial volume (%ICV), a surrogate for maximal premorbid total brain volume.

We assess trial designs using extensions of previous methodology<sup>16, 17</sup>. This involves building linear mixed models describing the pattern of the repeated measures, and then using these models to predict sample size requirements for different designs. We restrict attention to two-arm parallel trials in which treatment is expected to reduce the rate of progression by a constant amount throughout follow-up. We start with simple trials with one pre-randomization visit and one at the end of follow-up and then i) investigate the effects of incorporating interim visits at which efficacy measures are also made, ii) make allowance for plausible missing data patterns and iii) consider the potential utilization of information on covariates such as CAG repeat length. In clinical trials covariates can be utilized for various reasons<sup>18</sup>, and in various ways. We here consider two: i) *stratified randomization coupled with covariate adjustment*, where randomization is adapted to balance factors thought to influence outcomes; and, ii) *sample enrichment*, which aims to increase statistical power by restricting trial entry to those for whom treatment effects are expected to be largest. We also allow for the fact that the same factor may

be used for both enrichment (limiting trial entry to values within a specified range), and for further stratification within the selected range.

Sample size calculations require specification of the anticipated treatment effect. This choice will not affect the relative merits of the designs considered (if one design requires 10% more patients than another for one effect it will also require 10% more patients for others), so when making comparisons the choice is to some extent arbitrary. Nonetheless, it is informative to present results for particular anticipated effects, so we give results for 20% and 40% reductions in the mean rate of change, choices that are typical of the effects sought in larger HD trials<sup>19-21</sup> (see discussion). For UHDRS-TMS and UHDRS-TFC we consider treatment effects that are simple percentage reductions in the mean rate of change. All of the other potential outcomes considered here change over time even in people without disease<sup>8</sup>. Brain volumes decrease due to normal ageing, whilst performance on motor and cognitive tasks improve due to practice effects. A 100% effective treatment would most plausibly result in the mean change being the same as that in healthy controls. Therefore for outcomes other than UHDRS-TMS and UHDRS-TFC, we (like others<sup>17, 22, 23</sup>) consider treatment effects that are 20 and 40% reductions in the excess mean rate of change (over and above that seen in healthy controls). We also make the commonly adopted assumption that variability will be unaltered by treatment such that variability in both arms will mimic that seen in the early-HD patients.

### ***Statistical methods***

The two-stage statistical approach builds on previously described principles<sup>16</sup>. It is described in detail in the statistical appendix, and summarized here. The first stage develops appropriate linear mixed models for each outcome. The second uses parameter estimates from these models to predict the expected mean and variance of observed treatment effects, and hence required sample sizes, for each design of interest.

For the non-imaging outcomes (UHDRS-TMS, UHDRS-TFC, SDMT and speeded tapping IOI-SD) the basic model is a standard random slopes model relating the outcome to time since randomization, with the addition of random site and site-by-visit interactions. UHDRS-TMS, UHDRS-TFC and the speeded tapping IOI-SD were all transformed, as described in the appendix,

so that residuals better approximated a normal distribution. Models were fitted in the early-HD patients and, for SDMT and speeded tapping, jointly with an analogous model in controls. To compute sample sizes for designs incorporating stratified randomization coupled with covariate adjustment, fixed main effects of age, CAG repeat length, disease burden (a combination of CAG repeat length and age<sup>24</sup>) and disease stage and their interactions with time were added. Participants were included in the analysis if at least two (*i.e.* at least one measure of change) of the four measurements were available.

For the imaging outcomes the basic models are extensions of the models recommended for the analysis of repeated "direct" measures of change<sup>25</sup>, with further extensions analogous to those for the non-imaging variables, except that main effects of baseline volume instead of disease stage were used. Participants were included if they had at least one measurement of change.

To perform sample size calculations for enriched designs, we additionally fitted all models in subsets formed by separately dichotomizing on CAG repeat length ( $>/\leq 43$ , the median), disease burden<sup>24</sup> ( $>/\leq 382.2$  (the median value)), disease stage (I/II) and baseline brain and caudate volumes as a percentage of ICV (each dichotomized at the median from the full early-HD cohort). In these calculations we continue to assume stratified randomization and statistical adjustment for the enrichment factor and other relevant covariates.

As in earlier applications<sup>26</sup> uncertainty in sample size calculations was estimated using the bootstrap<sup>27</sup>. A pattern-mixture was used to evaluate the impact of dropouts and missing data<sup>28</sup>,<sup>29</sup>, with assumptions concerning these based on the patterns seen in TRACK-HD.

We present required sample sizes for 90% power to detect 20 and 40% treatment effects (only 20% in the appendices). Our results can be converted to required sample sizes for any other treatment effect since sample sizes are inversely proportional to the square of the treatment effect: *e.g.* switching from a 20% to a 25% reduction multiplies all sample sizes by 0.64 ( $(20/25)^2$ ). Analogously, switching from 90% to 80% power multiplies all sample sizes by 0.747 (see appendix).

## **Results:**

We analysed data from 123 early-HD patients and 123 controls (Table 1). For the non-imaging variables around three-quarters of the participants provided information at all visits, with less than 10% providing no data (through having measures at only one visit and hence no information on change over time). For the imaging variables just over 50% have measures of change from baseline to all of the follow-up visits, with around 10% of controls and 15% of early-HD patients providing no change measures (Supplementary Table 1).

Required sample sizes differ markedly between the imaging and non-imaging outcomes (Table 2). For the clinical outcomes (UHDRS-TMS and TFC) a one year trial with 90% statistical power to detect a 20% treatment effect requires of the order of 6000 (3000 per arm) patients (for a design that only stratifies by, and adjusts for, site in the stratified randomization and statistical analysis). For the imaging outcomes the analogous numbers are of the order of 1000, whilst for SDMT around 3000 patients are required. The sample size is also highly dependent on the assumed treatment effect: doubling the effect reducing the required sample size by a factor of four.

Other than the assumed treatment effect, for each outcome the primary driver of required sample size is follow-up duration. Increasing follow-up from one to two years has a greater effect than the further increase to three years (Table 2). This is because for all measures the standard deviation of the annualized rate of change declines to an asymptote (representing the between-subject standard deviation) as follow-up increases, with the steepness of the decline progressively decreasing (Figure 1).

The impact of adding interim visits at which efficacy measurements are made is much less marked. Adding one visit mid-way through follow-up has little impact on required sample sizes for the one- and two-year designs, but six-monthly interim visits do somewhat reduce requirements for the two and three-year designs. For a two-year clinical trial with six-monthly interim visits the most statistically efficient outcome was caudate atrophy expressed as a percentage of baseline volume, with 298 participants (149 per arm (95% CI (117, 269))) required for 90% statistical power to detect a 20% reduction. The analogous numbers for SDMT (the best



non-imaging outcome) were 412 (272, 700), with those for the other non-imaging outcomes being roughly double this (Table 2).

Adopting a design with a stratified randomization that incorporates factors other than site, and carrying out a statistical analysis that adjusts for these factors, can reduce sample size requirements (see Table 3 for two-year designs with six-monthly interim visits, supplementary table 2 for other designs). Reductions are fairly modest for the non-imaging outcomes; for a two-year clinical trial with six-monthly interim visits stratifying by (and adjusting for) age, CAG repeat length, disease burden and disease stage in addition to site only reduces samples sizes by between 5 and 15% (Table 3, scenarios A. and B.). However, for caudate atrophy gains are larger. For this outcome required sample sizes for a two-year trial with six-monthly interim visits are estimated to be smallest when atrophy is expressed as a percentage of baseline; here stratifying by and adjusting for age, CAG repeat length, disease burden and baseline caudate volume reduces the required sample size by a substantial 42% (Table 3, scenarios A. and B.).

Using inclusion/exclusion criteria for enrichment can also reduce sample size requirements (see Table 3, scenarios D. to F. for two-year designs with six-monthly interim visits, supplementary tables 3-5 for other designs). For example, for caudate atrophy (analyzed as a percentage of baseline) estimated sample sizes are approximately halved by restricting to patients with CAG repeat lengths of 44 and above. For SDMT, restriction to either those with greatest disease burden, or to stage II patients, reduces sample size requirements by around a third. However, it should be noted that in our exploration of enrichment, confidence intervals are wider than in earlier analyses because of the reduced numbers of patients considered. This imprecision is illustrated by the fact that for the UHDRS measures restriction to stage II patients actually increases estimated requirements.

Sample size calculations should always take into account dropouts and other anticipated missing data. More missing data is expected for imaging variables (through inadequate scans and because some are unable to tolerate imaging). The allowance that we make for missing data (table 3 scenario C. for two-year designs with six-monthly interim visits, supplementary table 6 for other designs) which base assumptions about missing data from TRACK-HD follow-up reflect this.

Including interim visits at which efficacy measures are made also has greater benefit when dropouts are anticipated. For example, for caudate atrophy (expressed as a percentage of baseline) adding six-monthly interim visits to a two-year trial reduces estimated sample size requirements by 19% (supplementary table 6) for a stratified design with dropouts compared with only 13% (supplementary table 2) when no dropouts are anticipated (and 9% (table 2) in a design only stratified by site).

Based on our analyses our general recommendation (figure 2, table 3 scenario C.) is for two-year clinical trials with six-monthly interim visits. With appropriate allowance for missing data, stratification and statistical adjustment (as described above), but no enrichment, a trial with 90% statistical power to detect a 20% treatment effect requires 107 (95% CI (90, 154)) patients per arm if caudate atrophy (expressed as a percentage of baseline) is the chosen outcome variable. With this design and SDMT as the outcome, sample sizes are multiplied by an approximate factor of four (410 (277, 712) per arm).

### **Discussion:**

Sample size and its corresponding resource requirements are important criteria when designing clinical trials. We here extend earlier methodology to compute required sample sizes for various HD clinical trial designs. Several interventions intended to slow the progression of HD have been tested in long-term clinical trials<sup>19-21, 30-32</sup>. None found statistically significant treatment effects. There are likely multiple reasons for these failures, including the absence of valid targets, pharmacokinetic limitations such as lack of penetration of the blood brain barrier, and insufficient statistical power.

Since no intervention in HD has been shown to be efficacious in slowing deterioration, defining clinically meaningful effects is problematic. Relative comparisons between required sample sizes for different designs are not contingent on the choice of treatment effect, but nonetheless it is useful to give context to comparisons by focusing on particular effects. Accordingly, we present results for 20% and 40% reductions in the mean rate of decline, reductions that are typical of those that the larger HD trials have attempted to detect. For example, the sample size

justification in the CARE trial<sup>20</sup> was based on an anticipated 40% reduction in the rate of TFC decline, whilst that for the Riluzole trial<sup>19</sup> was based on a mean 1 point change in TFC, which corresponds to an approximate 40% reduction (the mean change in the trial was around 2.4), whilst that for the coenzyme Q10 trial<sup>21</sup> was also based on a mean 1 point change in TFC, approximately 20% of the mean change of 4.5 seen during follow-up. Other HD trials<sup>31, 32</sup> have based their calculations on much larger, arguably over-optimistic, effects. Effects of between 20% and 40% have also been sought in other neurological diseases. For example, the successful MS-STAT trial<sup>33</sup> in multiple sclerosis was powered to detect a 25% reduction in whole brain atrophy. A review of trials in Alzheimer's disease<sup>23</sup> also considered 25% reductions (and 50% reductions in excess rate over and above that seen in people with disease). Reductions of between 20% and 40% represent, in our view, a clinically meaningful slowing of decline that can potentially be demonstrated in a trial of reasonable size and duration.

The size of the anticipated treatment effect has a strong, well-known impact on sample size; treatments anticipated to be twice as effective require only 25% of the participants. The relative merits of the various designs that we consider are unaffected by our choice of anticipated treatment effect, or by our choice of statistical power: varying these simply multiplies the required sample sizes for each design by the same factor. However, although relative impact is not affected by such considerations, the feasibility of a given design does depend on whether the sample size is achievable.

Our findings show that for trials where the outcome is a rate of change, the primary driver of sample size (other than the anticipated treatment effect) is length of follow-up. Larger gains in efficiency come from extending follow-up from one to two years compared with a further extension to three years. Increasing the number of interim visits has less impact: however, their value increases when allowance is made for anticipated dropouts and missing values. Due to space limitations, we have considered only regularly spaced interim visits here. As in other applications<sup>17</sup> our methodology can potentially be used to investigate the benefits of more frequent, or irregularly spaced, interim visits, but interpretation must be cautious when source data was not collected at such frequencies. When visits are frequent within-subject correlation

may differ in ways that are difficult to predict confidently. Inclusion of interim visits does increase costs and burden on subjects, especially in studies involving imaging. However, since some investigational products may require more frequent visits to evaluate safety, additional efficacy outcome measurements could be collected concomitantly, so reducing costs. Also, including interim assessments can support design decisions about future trials.

Stratification by factors such as age, CAG, disease burden, disease stage and baseline volume measures can reduce sample sizes, if coupled with appropriate statistical analysis. Our results potentially slightly over-estimate the benefits of such strategies since we assume that perfect balance on multiple stratification factors can be achieved, which may in practice not be possible, particularly when sample sizes are small and/or large numbers of sites are involved. Some stratification factors may then have to be omitted or minimization<sup>34</sup> considered as an alternative to stratified randomization. Enrichment is more contentious. It is possible that use of some treatments may ultimately be restricted to patients with higher CAG lengths and/or that enrichment may be necessary for targeted therapies that are only anticipated to work in a restricted population. In our analysis enrichment by restriction to patients with CAG repeat lengths of 44 and above reduced estimated sample sizes by up to 50%. However, it also reduced the available population by around half. Further, our calculations assume that the percentage reduction in rates of change achievable through treatment is unaltered by enrichment, implying that absolute reductions increase, which may be unrealistic.

Our most statistically efficient outcome was caudate atrophy expressed as a percentage of baseline volume. Without stratification or enrichment, 330 participants (165 per arm (95% CI (130, 280))) are required for 90% statistical power to detect a 20% reduction in caudate atrophy rates in a two-year follow-up trial with no interim visits. This required sample size can be reduced to an estimated 87 per arm (95% CI (73, 125)) in a design that scans participants every six months and stratifies by age, CAG repeat length and baseline caudate volume: although a plausible assumption concerning missing values gives a required sample size of 107 (95% CI (90, 154)) per arm. Of the non-imaging outcomes considered, the SDMT was most efficient (at least when practice effects were allowed for).

Choice of the most appropriate primary outcome measure for a particular intervention cannot be based only on sample size; biological plausibility, the expected effects of the intervention, the stage of the disease and the regulatory path forward all need consideration. We cannot assume that an intervention targeting the physiopathology of the disease will equally and simultaneously affect all outcomes. For example, a therapy aimed at maintaining white matter may not have a measurable effect on striatum structure. Likewise, SDMT but not UHDRS-TMS, is likely to capture an intervention targeting cognition. It is also possible that different effects will occur in different timeframes, these sometimes being unexpected. For example, clinical trials that tested anti-amyloid interventions in Alzheimer's disease showed an increase in brain atrophy, when a decrease was expected; despite some cognitive improvement and pharmacodynamic proof of amyloid reduction by PET scan<sup>35, 36</sup>. Accordingly trials should include a range of secondary outcomes and have built-in learning opportunities for the validation of others. Ideally, clinical trials should favor outcomes that best represent the disease from patients' perspectives. Unfortunately those that perhaps best fit this criterion are UHDRS-TMS and UHDRS-TFC, which we found to be the least statistically efficient. Further, both these scales may be even less efficient in a trial, since the increases in scores seen in TRACK-HD might in part be an artifact, arising from observers' expectation of deterioration without treatment. In trials they may also be affected by placebo effects arising from expectations of benefit. Structural Imaging outcomes are the most efficient, but they raise complex problems of translation to clinical relevance: so may be most appropriate in "proof of concept" phase II trials.

One limitation of our analysis is that we used a single naturalistic study conducted in just four highly specialized sites. Although absolute levels of the outcome variables differed between sites, rates of change did not differ markedly. The statistical models that we used did allow for variability in changes over time between sites through the inclusion of site-by-visit interaction terms, but they did not allow the anticipated treatment effect on rate of change to differ across sites. This means that our results can only be cautiously generalized to multi-centre studies with less specialized sites, where greater heterogeneity might be anticipated. Additionally, there is some evidence<sup>37</sup> that dropout rates in clinical trials may be higher than observed in TRACK-HD. Also, we do not have information on possible placebo effects which could be important,

Design optimization for clinical trials in early stage manifest Huntington's Disease

particularly for non-imaging outcomes<sup>38</sup>. Despite these limitations, we believe our work provides a useful contribution to the design of trials in early-HD. It provides a clear indication of the best duration, shows the benefits of stratified randomization when coupled with appropriate statistical analysis, and gives a firm basis for the choice of analytical method for imaging outcomes.

**TRACK-HD Investigators:**

	<b>Last name</b>	<b>First name</b>	<b>Location</b>
1.	Bechtel	Natalie	GHI, Muenster
2.	van den Bogaard	Simon J.A.	LUMC, Leiden
3.	Bohlen	Stefan	GHI, Muenster
4.	Campbell	Melissa	Monash
5.	Coleman	Allison	UBC, Vancouver
6.	Craufurd	David	Manchester
7.	Crawford	Helen	UCL
8.	Dar Santos	Rachelle	UBC, Vancouver
9.	Decolongon	Joji	UBC, Vancouver
10.	Dumas	Eve M.	LUMC, Leiden
11.	Hobbs	Nicola Z	UCL
12.	't Hart	Ellen P.	LUMC, Leiden
13.	Jauffret	Celine	Paris
14.	Justo	Damian	Paris
15.	Landwehrmeyer	Bernhard	Ulm
16.	Lahiri	Nayana	UCL
17.	Labuschagne	Izelle	Monash
18.	Nigaud	Kevin	Paris
19.	Read	Joy	UCL
20.	Say	Miranda J	UCL

21.	Sturrock	Aaron	UBC, Vancouver
-----	----------	-------	----------------

**Acknowledgments:**

TRACK-HD is supported by the CHDI Foundation, Inc., a not for profit organization dedicated to finding treatments for Huntington's disease. The authors wish to extend their gratitude to the TRACK-HD study participants and their families. Some of this work was undertaken at UCLH/UCL who acknowledge support from the Department of Health's NIHR Biomedical Research Centres.

**Author Roles:**

Chris Frost: Conception, design and execution of statistical analysis. Wrote first draft of manuscript.

Amy Mulick: Execution of statistical analysis. Review and critique of manuscript.

Rachael I. Scahill: Execution of research project. Review and critique of manuscript.

Gail Owen: Organization of research project. Review and critique of manuscript.

Elizabeth Aylward: Input into design of statistical analysis. Review and critique of manuscript.

Blair R. Leavitt: Execution of research project. Review and critique of manuscript.

Alexandra Durr: Execution of research project. Review and critique of manuscript.

Raymund A. C. Roos: Execution of research project. Review and critique of manuscript.

Beth Borowsky: Conception, design and organization of research project. Review and critique of manuscript.

Design optimization for clinical trials in early stage manifest Huntington's Disease

Julie C. Stout: Execution of research project. Review and critique of manuscript.

Ralf Reilmann: Execution of research project. Review and critique of manuscript.

Douglas R. Langbehn: Input into design of statistical analysis. Review and critique of manuscript.

Sarah J. Tabrizi: Conception, design, organization and execution of Research Project. Review and critique of manuscript.

Cristina Sampaio: Writing of sections of manuscript. Review and critique of manuscript.

**Financial Disclosures:**

Chris Frost received grants from NIHR, the Economic and Social Research Council, Effective Intervention, the Multiple Sclerosis Trials Collaboration and the Institute of Neurology, UCL. He received payment for consultancy work from CSL Behring. He received payment for his role as an external examiner for the University of Oxford.

Amy Mulick has no financial disclosures.

Rachael I. Scahill has no financial disclosures.

Gail Owen has no financial disclosures.

Elizabeth Aylward received grants from NIH.



Design optimization for clinical trials in early stage manifest Huntington's Disease

Blair R. Leavitt reports research grants from Teva, CHDI, and Lifemax. Personal Consulting fees from Roche, uniQure, Novartis, Lifemax and Raptor.

Alexandra Durr received research grants from the French Agency for Research, French Ministry for Social affairs and Health, Pfizer Inc. and Annapurna Therapeutics. She holds partly a patent B 06291873.5 on Anaplerotic therapy of Huntington's disease and other polyglutamine diseases.

Raymund A. C. Roos received grants from The Gossweiler Foundation and TEVA, both via the LUMC administration. He is an advisor for UniQure.

Beth Borowsky was an employee of CHDI Foundation, Inc, and is currently an employee of Teva Pharmaceuticals.

Julie C. Stout has served on an advisory board for Roche, has consulted to Prana Biotechnology, is treasurer and board member for the Huntington's Study Group, Inc., and conducts business implementing cognitive assessments at Stout Neuropsych Pty Ltd, with contracts from Teva, Vaccinex, Omeros, and Ionis.

Ralf Reilmann is founding director and owner of the George-Huntington-Institute, a private research institute focused on clinical and preclinical research in Huntington's disease, and QuantiMedis, a clinical research organization providing Q-Motor (quantitative motor) services in clinical trials and research. He holds appointments at the Dept. of Radiology of the University of Muenster and at the Department of Neurodegenerative Diseases and Hertie-Institute for Clinical Brain Research, University of Tuebingen. He provided consulting services, advisory board functions, clinical trial services, quantitative motor analyses, and/or lectures for Teva, Pfizer, uniQure, Ipsen, Vaccinex, Novartis, Raptor, Omeros, Siena Biotech, Neurosearch Inc., Lundbeck, Medivation, Wyeth, ISIS Pharma, Link Medicine, Prana Biotechnology, MEDA Pharma, Temmler Pharma, Desitin, AOP Orphan, and the Cure Huntington's Disease Initiative Foundation. He received grant support from the High-Q-Foundation, the Cure Huntington's

Design optimization for clinical trials in early stage manifest Huntington's Disease

Disease Initiative Foundation (CHDI), the Deutsche Forschungsgemeinschaft (DFG), the European Union (EU-FP7 program), the Bundesministerium für Bildung und Forschung (BMBF), the Deutsches Zentrum für Neurodegeneration und Entzündung (DZNE), and the European Huntington's Disease Network (EHDN).

Douglas R. Langbehn received funding for his work via a subcontract from UCL. The funds originated from the CHDI foundation. Dr. Langbehn also received additional funding from the CHDI foundation during this time. He also serves as a paid statistical consultant for Roche, Voyager, and Teva pharmaceutical companies for the design of HD trials.

Sarah J Tabrizi receives grant funding for her research from Medical Research Council UK, the Wellcome Trust, the EU FP7 Health Call, the Rosetrees Trust, Takeda Pharmaceuticals, NIHR UCL BRC, North Thames Local Clinical Research Network, Wolfson Foundation for Neurodegeneration and the CHDI Foundation. In the past three years Sarah has been on advisory boards or had consultancies with GSK, Ionis Pharmaceuticals (previously ISIS Pharmaceuticals), F. Hoffmann-La Roche Ltd, Ixico Technologies, Shire Human Genetic Therapies, Takeda Pharmaceuticals International and TEVA Pharmaceuticals. All honoraria for these consultancies and advisory boards were paid to UCL, Sarah's employer. Sarah has provided one-off consultancy services for Astex Pharmaceuticals, Optio Biopharma Solutions, GLG, Putnam Associates and All Global, for which all honoraria was paid to her directly. Through the offices of UCL Consultants Ltd, a wholly owned subsidiary of University College London, Sarah has undertaken consultancy services for F. Hoffmann-La Roche Ltd, GSK and TEVA Pharmaceuticals.

Christina Sampaio is an employee of CHDI Management Inc/ CHDI foundation. CHDI received consultancy fees in the last 12 months from Sealth, vtv Therapeutics and Nestle. She also received an Honorarium from IPMDS.

**References:**

1. Mestre T, Ferreira J, Coelho MM, Rosa M, Sampaio C. Therapeutic interventions for disease progression in Huntington's disease. *Cochrane Database Syst Rev*, 2009 Jul 8;(3):CD006456. doi: 10.1002/14651858.CD006456.pub2.
2. Sampaio C, Borowsky B, Reilmann R. Clinical trials in Huntington's disease: Interventions in early clinical development and newer methodological approaches. *Mov Disord* 2014; 29: 1419-1428.
3. Burman CF, Wiklund SJ. Modelling and simulation in the pharmaceutical industry—some reflections. *Pharmaceutical Statistics* 2011; 10: 508-516.
4. Critical Path Institute. U.S. Food and Drug Administration and European Medicines Agency Reach Landmark Decisions on Critical Path Institute's Clinical Trial Simulation Tool for Alzheimer's Disease. 2013, consulted December 2014: <http://c-path.org/wp-content/uploads/2014/03/US-FDA-EMA-agency-reach-landmark-decisions-C-Path-clinical-trial-simulation-tool-for-alzheimers-disease.pdf>.
5. Tabrizi SJ, Langbehn DR, Leavitt BR et. al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurology* 2009; 8: 791-801.
6. Tabrizi SJ, Scahill RI, Durr A et. al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. *Lancet Neurology* 2011; 10: 31-42.
7. Tabrizi SJ, Reilmann R, Roos RAC et. al. The TRACK-HD battery – potential endpoints for clinical trials in early stage Huntington's disease and insights into future trials for premanifest HD. *Lancet Neurology* 2012; 11: 42-53.
8. Tabrizi SJ, Scahill RI, Owen G et. al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurology* 2013; 12: 637-649.
9. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer, 2000.

10. Franciosi S, Shim Y, Lau M, Hayden MR, Leavitt BR. A systematic review and meta-analysis of clinical variables used in Huntington disease research *Movement Disorders* 2013; 28: 1987-1994.
11. Unified Huntington's disease rating scale: Reliability and consistency. *Movt Disord* 1996; 11: 136-142.
12. Huntington Study Group. Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 1996; 11: 136-142.
13. Bechtel N, Scahill RI, Rosas HD, Acharya T, van den Bogaard SJ, Jauffret C, Say MJ, Sturrock A, Johnson H, Onorato CE, Salat DH, Durr A, Leavitt BR, Roos RA, Landwehrmeyer GB, Langbehn DR, Stout JC, Tabrizi SJ, Reilmann R. Tapping linked to function and structure in premanifest and symptomatic Huntington disease. *Neurology* 2010; 75: 2150-2160.
14. Freeborough P, Fox NC. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans. Med. Imaging* 1997; 16: 623-9.
15. Aylward EH, et al., Longitudinal change in regional brain volumes in prodromal Huntington disease. *Journal of Neurology, Neurosurgery & Psychiatry* 2011; 82: 405-410.
16. Frost C, Kenward MG, Fox NC. Optimizing the design of clinical trials where the outcome is a rate. Can estimating a baseline rate in a run-in period increase efficiency? *Statistics in Medicine* 2008; 27: 3717-3731.
17. Schott JM, Frost C, Whitwell JL, MacManus DG, Boyes RG, Rossor MN, Fox NC. Combining short interval MRI in Alzheimer's disease: Implications for therapeutic trials. *Journal of Neurology* 2006; 253: 1147-1153.
18. Pocock SJ, Assmann SF, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine* 2002; 21:2917-30.
19. Landwehrmeyer GB, Dubois B, Garcia de Ye'benes J, Kremer B, Gaus W Kraus PH, Przuntek H, Dib M, Doble A, Fischer W, Ludolph AC and the European Huntington's Disease Initiative Study Group. Riluzole in Huntington's Disease: A 3-Year Randomized Controlled Study

- Annals of Neurology 2007; 62: 262–272.
20. The Huntington Study Group. A randomized, placebo-controlled trial of coenzyme Q10 and remacemide in Huntington's disease. *Neurology* 2001; 57:397–404.
  21. McGarry A, et. al.. A randomized, double-blind, placebo-controlled trial of coenzyme Q10 in Huntington disease. *Neurology*. 2017; 88:152-9.
  22. Cash DM, et. al.. Assessing atrophy measurement techniques in dementia: Results from the MIRIAD Atrophy Challenge. *NeuroImage*, 123: 149-164, 2015.
  23. Ard MC, Edland SD. Power Calculations for Clinical Trials in Alzheimer's Disease. *Journal of Alzheimer's Disease* 2011; 26:369–377.
  24. Penney JB Jr, Vonsattel JP, MacDonald ME, Gusella JF, Myers RH. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann Neurol* 1997; 41: 689–92.
  25. Frost C, Kenward MG, Fox NC. The analysis of repeated 'direct' measures of change illustrated with an application in longitudinal imaging. *Statistics in Medicine* 2004; 23: 3275-3286.
  26. Schott JM, Price SL, Frost C, Whitwell JL, Rossor MN, Fox NC. Measuring atrophy in Alzheimer's disease: a serial MRI study over 6 and 12 months. *Neurology* 2005; 65: 119-124.
  27. Efron B, Tibshirani RJ (1993). *An introduction to the bootstrap*, Chapman & Hall.
  28. Dawson JD, Lagakos SW. Analyzing laboratory marker changes in AIDS clinical trials. *Journal of AIDS* 1991; 4: 667-676.
  29. Dawson JD, Lagakos SW. Size and power of two-sample tests of repeated measures data. *Biometrics* 1993; 49: 1022-1032.
  30. Mestre T, Ferreira J, Coelho MM, Rosa M, Sampaio C. Therapeutic interventions for disease progression in Huntington's disease. *Cochrane Database of Systematic Reviews* 2009, Issue 3. Art. No.: CD006455. DOI: 10.1002/14651858.CD006455.pub2.
  31. Puri BK, Leavitt BR, Hayden MR, Ross CA, Rosenblatt A, Greenamyre JT, Hersch, S, Vaddadi KS, Sword A, Horrobin DF, Manku M, Murck H. Ethyl-EPA in Huntington disease. *Neurology* 2005; 65:286–292.

32. Peyser CE, Foistein M, Chase GA, Starkstem S, Brandt J, Cockrell JR, Byisma F, Coyle JT, McHugh PR, Foistein, SE. Trial of d-alpha-Tocopherol in Huntington's Disease. *American Journal of Psychiatry* 1995; 152: 1771-1775.
33. Chataway J, Schuerer N, Alsanousi A, Chan D, MacManus D, Hunter K, Anderson V, Bangham CRM, Clegg S, Nielsen C, Fox NC, Wilkie D, Nicholas JM, Calder VL, Greenwood J, Frost C, Nicholas R. Effect of high-dose simvastatin on brain atrophy and disability in secondary progressive multiple sclerosis (MS-STAT): a randomised, placebo-controlled, phase 2 trial. *Lancet* 2014; 383:2213-2221.
34. Pocock S, Simon R. Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial. *Biometrics* 1975; 31: 103–115.
35. Delrieu J, Ousset PJ, Caillaud C, Vellas B. 'Clinical trials in Alzheimer's disease': immunotherapy approaches. *Journal of Neurochemistry* 2012; 120: 186-193.
36. Fox NC et. al. Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology* 2005; 64: 1563-1572.
37. Borowsky B, Sampaio C. Experimental Therapeutics: moving forward in clinical trials, in *Huntington's Disease*, G. Bates, S. Tabrizi, and L.R. Jones, Editors. 2014, Oxford University Press: Oxford Monographs on Medical Genetics.
38. Murray D, Stoessl AJ. Mechanisms and therapeutic implications of the placebo effect in neurological and psychiatric conditions. *Pharmacology & Therapeutics* 2013; 140: 306-318.
39. Dawson DJ. Sample Sizes Based on Slopes and Other Summary Statistics. *Biometrics* 1998; 54: 323-330.